Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials

Thomas F. Willems,[1] Chris H. Rycroft,[1,2,*] , Michaeel Kazi,[1,2]

Juan C. Meza[1], Maciej Haranczyk[1,†]

[1] *Computational Research Division, Lawrence Berkeley National Laboratory, One Cyclotron Road, Mail Stop 50F-1650, Berkeley, CA 94720-8139, USA*

[2] *Department of Mathematics, University of California, Berkeley, CA 94720, USA*

Abstract:

Crystalline porous materials have a variety of uses, such as for catalysis and separation. Identifying suitable materials for a given application can, in principle, be done by screening material databases. Such a screening requires automated high-throughput analysis tools that calculate structural properties for all materials contained in a database so they can be compared with search queries, grouped or classified. One important aspect of the structural analysis of materials such as zeolites and metal organic frameworks is the investigation of the geometrical parameters describing pores. Here, we present algorithms and tools to efficiently calculate some of these important parameters. Our tools are based on the Voronoi decomposition, which for a given arrangement of atoms in a periodic domain provides a graph representation of the void space. The resulting Voronoi network is analyzed to obtain the diameter of the largest included sphere and the largest free sphere, which are two geometrical parameters that are frequently used to describe pore geometry. Accessibility of nodes in the network is also determined for a given

[*] Email: chr@math.berkeley.edu

[†] Corresponding author. Email: mharanczyk@lbl.gov; Fax: +1 (510) 486 5812

guest molecule and the resulting information is later used to retrieve dimensionality of channel systems as well as in Monte Carlo sampling of accessible surfaces and volumes. The presented algorithms are implemented in a software tool, Zeo++, which includes a modified version of the Voro++ library. We present example applications of our algorithms and tools using zeolite frameworks currently listed in the Atlas of Zeolite Frameworks.

Keywords:
Largest free sphere, largest included sphere, accessible surface area, accessible volume, channel dimensionality

1 Introduction

Porous materials contain complex networks of void channels and cages that are exploited in many industrial applications. The zeolite class of these materials is the most well-known as they have found wide use in industry since the late 1950s, with common applications as chemical catalysts and membranes for separations and water softeners [1,2,3,4]. There is increasing interest in utilizing zeolites as membranes or adsorbents for CO_2 capture applications [3]. In addition to zeolites, metal organic frameworks (MOFs) [5,6] and zeolitic imidazolate frameworks (ZIFs) [7] have recently generated interest for their potential use in gas separation or storage [8,9]. A key requirement for the success of any nanoporous material is that the chemical composition and pore topology must be optimal at the given conditions for a particular application. However, finding the optimal material is an arduous task, since the number of possible pore topologies is extremely large. There are approximately 190 unique zeolite frameworks known to exist today in more than 1400 zeolite crystals of various chemical composition and different geometrical parameters (See Ref. 10). However, these experimentally known zeolites constitute only a very small fraction of more than 2.7 million structures that are feasible on theoretical grounds [11,12], of which between 314,000 and 585,000 are predicted to be thermodynamically accessible as aluminosilicates, with the remainder potentially accessible via elemental substitution [13,14]. Databases of similar or greater magnitude can be developed for other nanoporous materials such as MOFs or ZIFs. As a result, new automated computational and chemoinformatic techniques need to be developed to characterize, categorize, and screen such large databases.

Recently, automated approaches capable of performing analysis of large sets of porous materials have started to emerge. For example, Blatov and coworkers have pursued the concept of natural tiling of periodic networks to find primitive building blocks in zeolites [15]. The group of Blaisten-Barojas have developed zeolite framework classifiers using a machine learning approach [16] while Foster *et al*. have presented a method to calculate geometrical parameters describing pores in zeolite materials [17]. The latter two approaches both make use of the Delaunay tessellation. For a given arrangement of atoms in a three-dimensional domain, the Delaunay tessellation is defined as the unique tetrahedral mesh on the atom positions such that the circumsphere of any tetrahedron contains no other atoms. Thus, the Delaunay tessellation identifies a set of spheres that occupy the voids within the structure. Delaunay tetrahedra and Delaunay circumspheres can be analyzed to obtain descriptors of the void space.

Foster *et al*. proposed an approach in which the circumspheres obtained from the Delaunay decomposition of a zeolite are analyzed to obtain two parameters frequently used to describe pore geometry in crystalline porous materials, namely the diameter of the largest included (D_i) and the largest free (d_f) spheres [18]. The largest included sphere points to the location of the largest cavity in a porous material and measures its size. In contrast, the largest free sphere corresponds to the largest spherical probe that can diffuse through a structure and measures a minimum restricting aperture on a diffusion path. Foster *et al*. noticed that the largest Delaunay circumsphere corresponds to the largest included sphere, whereas the largest free sphere can be obtained by investigating the Delaunay circumspheres obtained for a zeolite structure. The analysis is as follows. The empty Delaunay circumspheres frequently overlap and the circular intersection defines the diameter of the restricting aperture connecting them. The

diameter of these apertures places upper bounds on the diameter of the largest spherical probe that can pass from one Delaunay circumsphere to the other. When contiguous spheres overlap, channel systems are formed. Analysis of such channels can, in principle, provide information on the largest free sphere for a given zeolite structure. However, the authors have not presented the details of the approach thus far.

A contrasting approach has been recently proposed by Haldoupis *et al*. [19]. They calculated the largest included and free sphere diameters of zeolites and MOFs using a grid representation of the crystallographic unit cell of a material. Here, each grid point is assigned a distance to the surface of the nearest atom. The radius of the largest included sphere is then equal to the maximum value over all grid points. The largest free sphere is calculated with the multiple-labeling algorithm of Hoshen and Kopelman [20], which identifies and connects clusters of neighboring grid points with the assigned distance values above a threshold corresponding to a probe radius. Using these methods, Haldoupis *et al*. analyzed a hypothetical zeolite database containing more than 250 000 structures [21] as well as hundreds of MOFs. They also demonstrated an extension of the approach, where the distance grid is substituted with an energy grid obtained for a given probe, and then the same algorithm is used to estimate the net activation energy for diffusion of a probe through a material. The biggest disadvantage of the approach of Haldoupis *et al*., which we will address in the current study, is a high computational cost associated with computing the grid, which amount of up to few minutes per material [19].

Other important geometrical parameters used to characterize geometry of crystalline porous materials are accessible surface area (ASA) and accessible volume (AV). The accessible surface area of a system (a network of atoms), originally defined by Lee and Richards [22],

represents the surface traced by the center of a spherical probe (with radius r_{probe}) as it is rolled along the atomic surface. Another parameter, accessible volume, can be similarly defined as the volume reachable by the center of the probe. Although these parameters only constitute two potential screening variables, they are of especially great interest when building tools for identifying optimal absorption materials. For example, Frost *et al*. [23], when studying hydrogen absorption in several metal organic frameworks (MOFs), found a strong correlation between the amount absorbed and the accessible surface area when operating at intermediate pressures. In addition, Frost *et al*. also demonstrated that at high pressures, the amount absorbed correlates strongly with free volume, a parameter with only slight definitional differences from accessible volume.

Most current methods for calculating accessible surface area utilize a Monte Carlo integration approach based on the work of Shrake and Rupley [24]. They typically proceed as follows: by sampling points on a sphere (of radius $r_{atom} + r_{probe}$) centered around each atom of the material, the ASA contribution of each sphere corresponds to the fraction of viable points times the sampling sphere's surface area. Depending on the exact method, different criteria for determining a sampled point's viability have been proposed. From a simple geometric perspective, a viable point should not lie within another sampling sphere. This criteria, utilized by Duren *et al*. [25], excludes surfaces inside other atoms and within channels that are too narrow to fit the probe. Another approach, employed by Do *et al*. [26], calculates the potential between the probe and the surrounding atomic network and classifies points with negative or zero potential as viable. Similarly, most current methods for calculating accessible volume

typically use a Monte Carlo approach in which the acceptance of sampled points depends on energy criteria.

These criteria for accepting sampled points do not account for additional conditions that may lead to the overestimation of ASA and AV. For example, if atoms of a material define a cavity linked to a single channel, although a probe may fit inside the cavity, the narrowness of the channel can prevent the probe from ever reaching it. Because sampled points within these types of inaccessible pockets do not overlap with other sampling spheres, another condition is required to exclude these points. The problem of detection of inaccessible pockets is also faced when running molecular simulations to study diffusion or adsorption phenomena in porous materials [27,28]. Inaccessible pockets need to be excluded (or accounted for otherwise) from regions where guest molecules are placed during such simulations in order to reproduce experimental results. We have recently proposed two methods of detection of inaccessible pockets. One approach is based on visual analysis of abstract representations of porous materials, chemical hieroglyphs [29], which highlight accessible regions of a material. Another approach, which involves a grid-based front propagation technique executed to segment out channels and inaccessible pockets of a material, can be fully automated [30,31]. In this case, each point of the grid holds information about whether the probe can reside in the corresponding volume element, which is determined using the geometry-based or energy-based criterion. In either case, generating the grid representation of a material is a costly procedure that is not always justified, especially when one is interested in obtaining geometry-based estimates of ASA and/or AV for later sets of materials. Therefore, there is a need for fast approaches that determine the accessibility of pockets and exploit the resulting information in calculations of ASA and AV.

In the current contribution, we present a set of algorithms to calculate important geometrical characteristics of crystalline porous materials such as zeolites. We explore the Voronoi decomposition as an efficient technique to provide basic geometrical characteristics of a porous material and its void space as well as the pore accessibility information to be included in ASA and AV calculations. When performing a Voronoi decomposition, the space surrounding *n* points is divided into *n* polyhedral cells such that each of their faces is a plane equidistant from the two points sharing the face. Edges of such cells overlap with lines equidistant to neighboring points (three points in a general asymmetric case), whereas vertices of cells, the Voronoi nodes, are equidistant from neighboring points (four points in a general asymmetric case). The Voronoi network, built of such nodes and edges, maps the void space surrounding the points. The Voronoi tessellation is the dual of Delaunay tessellation, as the centers of Delaunay circumspheres correspond to nodes in Voronoi network. Analysis of such a network is fairly straightforward and can provide parameters such as the diameter of the largest free sphere as well as more detailed information about void space geometry and topology. The Voronoi decomposition has already been used in the analysis of crystalline materials [32] and their voids [33] as well as membranes [34] and has been suggested as a tool to investigate ion transport pathways in crystals [35].

Our implementation is based on a modification to the VORO++ software library developed by one of us [36,37].The resulting tool provides the Voronoi network supplemented with information relevant to analysis of the void space in porous materials, such as the maximum diameters of spherical probes that can travel along the edges of the network. Our modifications to the original version of VORO++ enable us to perform the Voronoi decomposition for periodic

systems with both rectangular and non-rectangular periodic unit cells. We present an efficient algorithm to calculate the diameters of the largest included and largest free spheres for a given structure. Moreover, we provide algorithms to determine accessibility of nodes of the Voronoi networks for a particular probe size as well as Monte Carlo integration of ASA and AV procedure that can use the resulting information. The discussed algorithms have been implemented in the Zeo++ package [38], which is demonstrated on a task of characterization of selected zeolites from the International Zeolite Association (IZA) database.

2 Methods

2.1 Calculation of the Voronoi network

The Voronoi network is computed using a modified version of VORO++, an open source library for three-dimensional Voronoi calculations [36,37]. The library is based upon individually computing the Voronoi cell associated with each atom, which is internally stored as a collection of edges and vertices. The details of our implementation of the Voronoi decomposition are provided in the Supporting Materials. An example computation of the Voronoi cells for the orthogonal case of EDI zeolite and non-orthogonal case of NPO zeolite are shown in Fig. 1.

During computation of the Voronoi network, various important parameters are tabulated for use in later analysis. In particular, for each edge and vertex, the minimum distance to an atom is stored. This distance can be defined as the distance to the atomic surface if radii of atoms are specified. Vertices are stored within the provided periodic unit cell, and for each edge, a periodic displacement vector (PDV) (i,j,k) is also stored that determines whether an edge connects two

vertices in a different periodic image. The *i*,*j*,*k* components of this vector point to the neighboring periodic cell in directions defined by the unit cell vector **a**, **b**, and **c**, respectively.

The library can also handle chemical systems with atoms of various radii, making it possible to study materials with chemical elements often encountered in MOFs or ZIFs. To do this, it makes use of the radical Voronoi tessellation in which the Voronoi cell of a particle *i* with position $\mathbf{x}_i$ and radius r_i is given by the volume satisfying

$$d(\mathbf{x},\mathbf{x}_i)^2 - r_i^2 < d(\mathbf{x},\mathbf{x}_j)^2 - r_j^2 \qquad (1)$$

for all other particles *j*. The square weighting ensures that the computed cells are convex polyhedra – without this, the cells have curved boundaries, which is more difficult to address computationally. The radical Voronoi tessellation is a common approach to handle polydisperse particle arrangements and has been previously shown to be a good method of constructing a network for porosity calculations of unequal spheres [39].

Apart from applications discussed in this study, our library can be used to build numerous other automated, high-throughput tools for analysis of void space in porous materials [40].

2.2 Analysis of the Voronoi network and characterization of a material

The obtained Voronoi network represents the void space in a porous material. Analysis of such a Voronoi network may provide information about the largest opening within the structure, the largest spherical probe that can traverse through the structure, topology of channel systems, and other parameters of interest. Here, we present how this information can be used to obtain structural descriptors such as the diameter of the largest included sphere (D_i), the diameter of the

largest free sphere (d_f), accessible surface area (ASA), accessible volume (AV) and the dimensionality of channel systems (dim).

2.2.1 The largest included and the largest free sphere

The largest included sphere for a given structure is simply the largest distance assigned to the Voronoi nodes. The algorithm iterates over all Voronoi nodes in a periodic unit cell of a structure and finds the node with the largest distance to a neighboring atom. .

Analysis of the Voronoi network also provides information about the size of the largest spherical probe that can travel within the void space. This analysis involves investigation of connectivity between Voronoi nodes. For example, finding the diameter of the largest spherical probe that can travel between two nodes involves finding the path in the Voronoi network that leads through the nodes and edges with largest distances to atoms (the path through the sections of the void space with the largest opening). We find such an optimal path using an implementation of Dijkstra's "lowest-cost path" algorithm [41]. Here, we relate the cost of accessing each node or passing through each edge with the corresponding distances to atoms assigned to the nodes and edges, such that the lowest cost is assigned to the largest opening in a structure.

The term 'largest free sphere' usually refers to the maximum size of a spherical probe that can travel through the void space in a structure by at least one periodic lattice translation. The void space may be shaped to form channels in many directions. Since crystalline porous materials form 3D networks built by replicating a periodic unit cell along each of three principal directions, **a**, **b**, and **c**, the travel path directions will always involve one or more of the principal directions. The largest free sphere for a given material can be defined as maximum of largest free

spheres calculated for three directions involving each of the principal directions (with corresponding diameters d_a, d_b and d_c, respectively). For instance, when considering traversal in the **a**-direction, the probe must pass through edges connecting the unit cell with its neighboring copies in the -**a** direction (the source edges with periodic displacement vector (-1,*,*), where * denotes any number), then through the source nodes (ones connected to source edges), then through other nodes and edges within the unit cell to finally reach "the sink nodes" and exit the unit cell through the edges pointing out to the neighboring unit cells in the +**a** direction (the sink edges with periodic displacement vector (1,*,*)). The largest probe that can travel along this path can be calculated from our adaptation of Dijkstra's algorithm. In particular, the largest spherical probe that can travel from source edges to sink edges is calculated using edges contained within the periodic unit cell and any edges that do not lead to a unit cell with a different **a**-coordinate (edges with displacement vectors (0,*,*)). In other words, periodicity is disabled within the direction of interest but enabled within the other two directions. In some materials, channels corresponding to largest free spheres are not perfectly aligned with principal directions of the periodic unit cell. In these cases, the corresponding largest free sphere can travel along all of the principal directions, a combination of which is required to define the direction of channel. Taking the maximum of these defined largest free spheres (d_a, d_b and d_c) as the definition of the largest free sphere for a material ensures that the latter gives the correct results despite the orientation of the travel path of the largest free sphere. Additionally, having identified the largest free sphere diameter, the largest included sphere diameter available to the largest free sphere can be calculated from a subset of the network available to this free sphere. Further details on implementation of the approach are presented in the Supporting Materials.

2.2.2 Determination of Node Accessibility

In order to calculate accessible surface area and accessible volume with exclusion of inaccessible pockets, the accessibility of the void space must be determined. In this study, the void space is represented by a Voronoi network in which each Voronoi node is classified as accessible (constituent of a channel) or inaccessible (constituent of an inaccessible pocket). Although the periodicity of porous systems typically complicates analysis of their underlying networks, it simplifies the determination of node accessibility. Because channels must propagate through an infinite number of unit cells, each of which contains a finite number of Voronoi nodes, channels must also be periodic. As a result, a node in a Voronoi network constitutes a channel (and is therefore accessible) only if it is connected, directly or indirectly, to its periodic image in another unit cell. Furthermore, a node connected to a channel must also be part of the channel. From these basic premises, an algorithm for identifying probe-dependent channels arises as follows:

1) Calculate the periodic Voronoi network for the unit cell of the porous material
2) Remove all nodes and edges in the Voronoi network for which the assigned distance to the nearest atomic surface is less than r_{probe} (as these would not allow the probe to pass).
3) Select an unvisited node as a starting point and record its periodic displacement vector (PDV) as $(0, 0, 0)$. Using the edges leading directly from this node, place the ids and PDVs of all directly connected nodes on a stack. While the stack contains nodes, remove the topmost node and perform the following analysis:
    i) If the node is unvisited, record its ID and PDV and add all of its direct neighbors and their corresponding PDVs to the stack.

ii) If the node was visited with a different PDV, all the connected nodes are accessible.

Once the stack is empty, if the nodes were not labeled as accessible, they are all inaccessible.

4) Repeat step 3 until all nodes have been visited.

Using this algorithm, each iteration of the 3rd step in which the nodes are accessible identifies the constituents of a channel.

2.2.3 Dimensionality of channel systems

Dimensionality of a channel system is another important parameter describing a material as it provides information about the number of directions in which guest molecules can diffuse. Obviously, definition of a channel (or a valid diffusion direction) requires setting a threshold value describing the minimal probe size that can traverse it. Typically, a water molecule of a diameter of ca 2.8Å is used. Dimensionality of a channel system is usually established by visual inspection of a structure, but in the following paragraph we discuss how analysis of Voronoi networks can be used to automatically determine dimensionality of channels systems.

Determination of channel dimensionality arises naturally from a simple extension to the node accessibility algorithm outlined in the previous section. Whenever a node is encountered in two different unit cells, a "loop" through a periodic boundary has been found. Such a loop means that a node is connected to its copy in another periodic box and the difference in PDVs indicates the direction of a connecting channel. Therefore, when algorithm 2.2.2 is run and a node's neighbors are added to the stack, a simple test should be performed to determine if any of the neighbors were already encountered with a different PDV. If so, the difference in PDVs should be recorded. For each channel that the algorithm identifies, the channel dimensionality (one, two or three-dimensional) can then be computed by determining the dimensionality of the subspace

spanned by the channel directions. If an iteration of step 3 does not identify any channel directions, the zero-dimensionality results from the existence of an inaccessible pocket.

2.2.4 Calculation of Accessible Surface Area and Accessible Volume

Having determined the accessibility of each Voronoi node, this information can be incorporated in an algorithm that calculates accessible surface area. The original ASA method utilizes a Monte Carlo integration approach where points are sampled on spheres (of radius $r_{atom} + r_{probe}$) centered on each atom. The ASA contribution of each sphere corresponds to the fraction of viable points times its surface area. Therefore, in order to exclude inaccessible pockets, the accessibility of Voronoi nodes must be extended to that of sampled points. Because Voronoi cells are convex polyhedra, a line segment can always be drawn from a Voronoi node to a point on a sampling sphere within the cell. If such a line segment only intersects the sampling sphere at the sampling point, the probe can move in a straight path from the sampled point to the Voronoi node. Because at least one such non-intersecting line segment exists, the accessibility of a point can always be determined by that of the surrounding nodes. Utilizing these key facts, an algorithm for determining accessible surface area is as follows:

I. Perform a radical Voronoi tessellation for a set of spheres centered on each atom, where the i^{th} sphere has radius $r_{atom\ i} + r_{probe}$.

II. Determine the accessibility of each Voronoi node for a probe of radius zero.

III. Randomly sample points on each sphere. For each sampled point:

   a. Determine whether the sampled point lies within another sampling sphere. If so, the probe would overlap with another atom and the sampled point is not viable.

b. If the point is still viable, determine in which Voronoi cell it lies and the cell's corresponding central atom. Draw line segments from the Voronoi nodes of this cell to the sampling point. Imagine a sphere centered on the central atom sized such that the sampled point lies on its surface. The accessibility of the first Voronoi node whose line segment only intersects this sphere at the sampling point determines the point's viability.

IV. Multiply the fraction of viable points of III by the surface area of the sampling sphere. This quantity represents the ASA contribution of the atom. Add it to the total accessible surface area.

One particularly important point is that the Voronoi decomposition uses the combined atomic and probe radii. This ensures that if a point within a Voronoi cell does not lie within the sampling sphere centered in the cell, it will be empty because of the d^2-r^2 criterion using during the radical tessellation.

Only a slight alteration of the technique used to calculate ASA is required to calculate probe-dependent accessible volume. While the ASA calculation sampled points on spheres centered on each atom, the accessible volume algorithm differs in that points are randomly sampled across the entire unit cell. After a point is sampled, the distance between it and every atomic surface must exceed r_{probe} to avoid probe-atom overlap. If this requirement is satisfied, the point's viability can be determined in the same manner as in the ASA calculation using Voronoi nodes. The accessible void volume is then equal to the product of the fraction of valid sampled points and the unit cell volume.

3 Results and discussions

To demonstrate the application of the tools discussed here, we have calculated the diameters of the largest included and the largest free sphere in known zeolite frameworks contained in the Atlas of Zeolite Framework types [42] from the International Zeolite Association (IZA). The current database contains 194 unique frameworks for which atomic coordinates are available. These structures are hypothetical materials generated by converting experimental structures into pure silica counterparts before applying geometric optimization using the DLS76 program [43]. In our calculations, following Ref. 17, we have assumed radii of 1.35Å for both O and Si atoms.

The obtained results for all 194 IZA zeolites are collected and presented in Table S-1 of Supporting Materials. Moreover, results for 29 structures not included in Ref. 17 are presented in Table 1. Our implementation of the free sphere calculation is fast and efficient. Typically, the analysis of one structure takes less than a second, and processing the entire database of 194 zeolites took 17 seconds on an 2.4GHz Intel i7 desktop.

To verify our algorithm for calculation of diameters of the largest included and free sphere, we compared our results with the results for 165 zeolites obtained previously by Foster *et al*. [17] and Haldoupis *et al*. [19]. As shown in Figure S-5, our results are in excellent agreement with previous studies. Largest included sphere diameters for all 165 zeolites match the results of Haldoupis *et al*., with differences lower than 0.1Å, corresponding to the grid spacing used in Ref. 19. For three structures, similarly to Ref. 19, we noticed deviations below 0.5Å, most likely due to slight differences in the crystal structures used in these sets of calculations. Similarly, the calculated free sphere diameters do not show important discrepancies when compared with the results of Haldoupis *et al*. However, for the zeolites AFN and RUT, our results were 0.5Å larger

than those of Foster *et al*. As already pointed by Haldoupis *et al*., in these zeolites the channels corresponding to the largest free sphere are not oriented along crystallographics axis and were not considered by Foster *et al*. Our free sphere algorithm calculates the diameter of the largest free sphere but it does not analyze the largest free sphere path. Therefore, the direction/orientation of the path and the corresponding channel is not available. However, as demonstrated by Haldoupis *et al*. in Refs. 19 and 21, the obtained information (the diameter of largest free sphere) is sufficient to select interesting materials from a database for a follow up characterization using molecular simulations methodology.

Real zeolites, either natural or synthetic, are not necessarily all-siliceous. They may also contain extra framework cations or other adsorbed phases which render changes in the crystal structure with respect to the ideal crystal. Our tool can be also used to compare different experimental and theoretical crystal structures. For example, our recent study [44], demonstrated how changes in the crystal structure, reflected by the changes of the largest free sphere diameter can influence adsorption and diffusion of methane in cage-type zeolites.

The implementation of accessible surface and accessible volume was first verified on example MOF materials included in the study of Duren et al. [25]. However, these materials do not have inaccessible pockets within the void space. Therefore, to apply the algorithms on a system with inaccessible pockets, the zeolite DDR was analyzed as depicted in Figure 2. Analysis of the accessibility of the Voronoi network for a probe of diameter of 3.2Å (representing a nitrogen molecule) revealed the existence of inaccessible nodes, resulting in the sampled surface area presented in Figure 2. As demonstrated in the figure, the DDR structure has two-dimensional channel systems spreading in both the **a** ([100]) and **b** ([010]) directions with a

corresponding largest free sphere diameter of 3.59Å. Two-dimensional channels are separated from each other along the **c** ([001]) direction and the intermediate space is filled with inaccessible pockets. The latter are connected to the channel system by channels of 2.57Å diameter, a value significantly lower than the selected threshold of 3.2Å. The calculation of accessible surface area and accessible volume takes on average only about 2 seconds. One can therefore use our tool to characterize large sets of materials with the desired sets of thresholds and parameters such as atomic and probe radii. Figure 3 presents histograms of calculated surface area and void volume for a set of 194 IZA zeolites using spherical probes of 1.6Å and 1.4Å radii, respectively. In each case, the calculations were preformed with and without exclusion of inaccessible pockets to obtain, respectively, accessible and total surface area and void volume. Generally, the number of zeolites assigned to particular bins differ in cases of total and accessible surface areas and void volumes. The latter fact clearly suggests that inaccessible pockets are common among IZA zeolites and therefore our (or a similar) approach should be used to correctly calculate ASA and AV.

To verify the channel dimensionality algorithm, we compared the results of our analysis with the information contained in the current online version of the IZA database. It is our understanding that the assignment of dimensionality of channels in the IZA database was performed manually by researchers and therefore does not use a precise threshold value describing a diameter of a valid channel. Typically, a water molecule is taken as a probe and a channel of diameter of roughly about 2.8Å is considered a valid channel. In our analysis, however, we have to use an exact number as a threshold value. In some cases, when the diameter of the largest channels in a particular direction is close to the threshold values, the qualitatively

different results can be obtained by slight variation of the threshold. For these reasons, we performed our analysis for a number of thresholds in the range of 2.6-3.0Å to find out if any of the obtained results matched the information of the IZA database. For only 7 out of 194 zeolites, our algorithm predicted different dimensionality than the database. In the case of three structures, ANA, CGF and IMF, our algorithm predicted no channels, 1-dimensional and 2-dimensional channels respectively, whereas IZA identified these as 3-, 2- and 3-dimensional. In these three cases, our algorithm underestimated dimensionality due to not taking into account channels of highly-non-circular cross-sections. Such channels cannot be taken into account in our model where we use a spherical probe model. In the case of four other structures, namely BOG, MON, OFF and SBE, our algorithm predicts 3-dimensional channel systems whereas IZA recognizes 2, 2, 1 and 2-dimensional channels respectively. To investigate these cases, we generated pore landscapes for these structures corresponding to the probe of 2.8Å diameter. As seen in Figure 4, the channel system clearly allows the probe to travel in three directions positively supporting our results.

Our dimensionality analysis algorithm provides an efficient and automatic way to determine dimensionality of channels systems, and in general, a way to classify large databases of materials on the basis of the dimensionality of channel systems. Our analysis does not take into account the shape of a channel, for example, a straight channel and zig-zag channel are indistinguishable. Similarly, information about the angles between channels is not taken into account. Our algorithms, however, could be further extended to take into account these circumstances. One can envision an additional step that takes into account total length of a

channel such that a zig-zag channel would be seen as a longer path than a straight path connecting two nodes of Voronoi network.

4. Discussion

An obvious deficiency of the geometrical structural descriptors described in this article is the approximation that a spherical probe models the traversing molecule. In practice, most molecules of interest, including common solvents or gases, rarely have a spherical shape. We have recently proposed an advanced approach [45], in which a spherical probe is replaced with one resembling the shape and flexibility of a "real" molecule – a complex object built from solid blocks connected by flexible links. Such advanced probes are able to change shape during the traversal of a porous material, reaching areas not accessible to either a single large spherical probe or rigid molecular-shaped probes. We presented the problem as an Eikonal equation in configuration space, in which the cost of entering each point in configuration space corresponds to the local geometry. Although this approach can give accurate predictions of possible diffusion pathways and accessible volumes, its computational cost is much greater.

The computational cost is one of the major factors differentiating grid-based and non-grid-based approaches to the analysis of porous materials. For example, Haldoupis *et al*. reports that the time required to calculate the largest free sphere diameter is up to few minutes. With our current approach, we observe calculation times of up to few seconds. Our approach is therefore better suited for pruning large databases at the early stage of screening when less accurate approaches, such as those that assume the spherical shape of a probe or ones based only on geometry considerations, are acceptable. The more expensive, grid-based approaches have an advantage that they can relatively easy modified to include energy considerations [19,30,45] or

complex non-spherical-shape probes [45]. They may be better suited for more accurate screening of preselected sets of materials. Moreover, the accuracy of grid-based methods is typically controlled by grid spacing and therefore, high accuracy may require significant resources. At the same time, grid-based approaches can easily take an advantage of novel computing platforms such General Purpose Graphical Processing Units (GPGPU) [31] as calculation of a grid representing a material or a configuration space of a guest inside a material is a naturally parallel task.

Finally, we would like to stress that although our approaches to the analysis of porous materials are automatic and can be used in high-throughput setups, they do not release users from the responsibility of curation of the input data. In particular, experimental crystal structures of porous materials often contain additional species such as water or other solvents. Unless specifically desired, these species have to be removed before using our tools in order to obtain uncompromised descriptors of a material's structure.

5 Conclusions

The Voronoi network obtained by performing a Voronoi decomposition on atoms contained in a periodic unit cell of a crystalline porous material maps the void space of the material. The obtained Voronoi network contains information about the topology and geometry of a porous material and can be studied to retrieve structural parameters, descriptors and fingerprints. Such a Voronoi network is generated by our implementation of the VORO++ library that facilitates work with crystalline materials. Our library can perform Voronoi decompositions for periodic systems with rectangular and non-rectangular periodic unit cells and can also account for atoms of different sizes. In addition, it is also robust and may facilitate development

of tools for automatic and high-throughput analysis of porous materials, which are required to process a large number of structures contained in material databases.

We have presented algorithms to calculate 1) the diameters of the largest included and largest free sphere, which are simple parameters describing geometry of the channel system in porous materials; 2) determine dimensionality of channel systems present in a material as well as 3) use information contained in the Voronoi network to calculate probe accessible surface and accessible volume. These algorithms have been implemented in the Zeo++ software tool, which is available to the community, together with its source code, on free of charge basis. We expect that many approaches for deriving structural information and descriptors from Voronoi network will emerge in the future enabling Qualitative Structure Property Relationships (QSPR) types of studies of porous materials.

Acknowledgements

The authors would like to thank Prof. Berend Smit for many fruitful discussions. This research was supported by the U. S. Department of Energy under contract DE-AC02-05CH11231. This work was also supported in part (to M. H. and T.F.W.) jointly by DOE Office of Basic Energy Sciences and the Office of Advanced Scientific Computing Research through SciDAC project #CSNEW918 entitled "Knowledge guided screening tools for identification of porous materials for CO_2 separations". J.C.M. and T.F.W were also supported as part of the Center for Gas Separations Relevant to Clean Energy Technologies, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy

Sciences under Award Number DE-SC0001015. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

[1] *Handbook of Zeolite Science and Technology*, edited by Auerbach, S.M.; Carrado, K.A.; Dutta, P. K. Marcel Dekker. New York, USA, 2004.

[2] B. Smit, T.L.M. Maesen, Nature 457 (2008) 671-677.

[3] B. Smit, T.L.M. Maesen, Chem. Rev. 108 (2008) 4125-4184.

[4] R. Krishna, J.M. van Baten, Chem. Eng. J. 133 (2007) 121-131.

[5] A.R. Millward, O.M. Yaghi, J. Am. Chem. Soc. 127 (2005) 17998-17999.

[6] K.S. Walton, A.R. Millward, D. Dubbeldam, H. Frost, J.J. Low, O.M. Yaghi, J. Am. Chem. Soc. 130 (2008) 406-407.

[7] R. Banerjee, A. Phan, B. Wang, C. Knobler, H. Furukawa, M. O'Keeffe, O.M. Yaghi, Science 319 (2008) 939-94.

[8] K. Sumida, M.R. Hill, S. Horike, A. Dailly, J.R. Long, J. Am. Chem. Soc. 131 (2009) 15120-15121.

[9] H.J. Choi, M. Dinca, J.R. Long, J. Am. Chem. Soc. 130 (2008) 7848-7850.

[10] (a) S. Yang, M. Lach-hab, I.I. Vaisman, E. Blaisten-Barojas, X. Li, V.L. Karen, J. Phys. Chem. Ref. Data 39 (2010) 033102-45; (b) S. Yang, M. Lach-hab, I. I. Vaisman, E. Estela Blaisten-Barojas, J. Phys. Chem. C 113 (2009) 21721-21725.

[11] M.D. Foster, M.M.J. Treacy, http://www.hypotheticalzeolites.net (accessed Nov 13, 2009).

[12] D.J. Earl, M.W. Deem, Ind. Eng. Chem. 45 (2006) 5449-5454.

[13] M.W. Deem, R. Pophale, P.A. Cheeseman, D.J. Earl, J. Phys. Chem. C 113 (2009) 21353–21360.

[14] R. Pophale, P.A. Cheeseman, M.W. Deem, Phys. Chem. Chem. Phys. DOI: 10.1039/c0cp02255a

[15] V.A. Blatov, O. Delgado-Friedrichs, M. O'Keeffe, D.M. Proserpio, Acta Cryst. A63 (2007) 418–425.

[16] a) M. Lach-hab, S. Yang, I. I. Vaisman , E. Blaisten-Barojas, Mol. Inform. 29 (2010) 297-301; b) D. A. Carr, M. Lach-hab, S. Yang, I. I. Vaisman, and E. Blaisten-Barojas, Micropor. Mesopor. Mater. 117 (2009) 339-349.

[17] M.D. Foster, I. Rivin, M.M.J. Treacy, O. Delgado, Micropor. Mesopor. Mater. 90 (2006) 32–38.

[18] H. Li, A. Laine, M. O'Keeffe, O.M. Yaghi, Science 283 (1999) 1145-1147.

[19] E. Haldoupis, S. Nair, D.S. Sholl, J. Am. Chem. Soc. 132 (2010) 7528–7539.

[20] J. Hoshen, R. Kopelman R., Phys. Rev. B 14 (1976) 3438-3445.

[21] E. Haldoupis, S. Nair, D.S. Sholl, Phys. Chem. Chem. Phys. 13 (2011) 5053-5060.

[22] B. Lee, F. M. Richards, J. Mol. Biol. 55 (1971) 379-490.

[23] H. Frost, T. Duren, R.Q. Snurr, J. Phys. Chem. B 110 (2006) 9565-9570.

[24] A. Shrake, J. A. Rupley, J. Mol. Biol. 79 (1973) 361-371.

[25] T. Düren, F. Millange, G. Férey, K.S. Walton, R.Q. Snurr, J. Phys. Chem. C 111 (2007) 15350-15356.

[26] D.D. Do, L.F. Herrera, H.D. Do, J. Colloid Interface Sci. 328 (2008) 110–119.

[27] (a) D. Dubbeldam,B. Smit, B., *J. Phys. Chem. B 107 (2003)* 12138-12152; (b) S.P. Bates, W.J.M. van Well, R.A. van Santen, B. Smit, *J. Am. Chem. Soc. 118 (1996)* 6753-6759.

[28] Krishna, R.; van Baten, J. M. Langmuir 26 (2010) 2975–2978.

[29] K. Theisen, B. Smit, M. Haranczyk, J. Chem. Inf. Model. 50 (2010) 461-469.

[30] M. Haranczyk, J.A. Sethian, J. Chem. Theory Comput. 6 (2010) 3472-3480.

[31] R.L. Martin, Prabhat, D.D. Donofrio, J.A. Sethian, M. Haranczyk, Int. J. High Perform. Comput. Appl. submitted

[32] V.A. Blatov, Cryst. Rev. 10 (2004), 249-318.

[33] V.A. Blatov, A.P. Shevchenko, Acta Cryst. A59 (2003) 34-44.

[34] M.G. Alinchenko, A.V. Anikeenko, N.N. Medvedev, V.P. Voloshin, M. Mezei, P. Jedlovszky, J. Phys. Chem. B 108 (2004) 19056-19067.

[35] V.A. Blatov, G.D. Ilyushin, O.A. Blatova, N.A. Anurova, A.K. Ivanov-Schits, L.N. Dem'yanets, Acta Cryst. B62 (2006) 1010-1018.

[36] a) Original library available at http://math.lbl.gov/voro++/ (June 1st 2010); b) The modified library discussed here is available from the authors upon request.

[37] C. H. Rycroft, Voro++: a three-dimensional Voronoi cell library in C++, Lawrence Berkeley National Laboratory, Paper LBNL-1430E, January 23rd 2009.

[38] Zeo++ is available from the authors upon request

[39] S.C. van der Marck, Phys. Rev. Lett. 77 (1996) 1785-1788.

[40] R.L Martin, M. Haranczyk, in preparation

[41] E.W. Dijkstra, Numerische Math. 1 (1959) 269–271.

[42] C. Baerlocher, W. M. Meier, D. H. Olson, Atlas of Zeolite Framework Types, seventh edition, Elsevier, Amsterdam, NL, 2007.

[43] C. Baerlocher, A. Hepp, W.M. Meier, DLS-76 – A program for simulation of crystal structures by geometric refinement, ETH Report, Zurich, 1977.

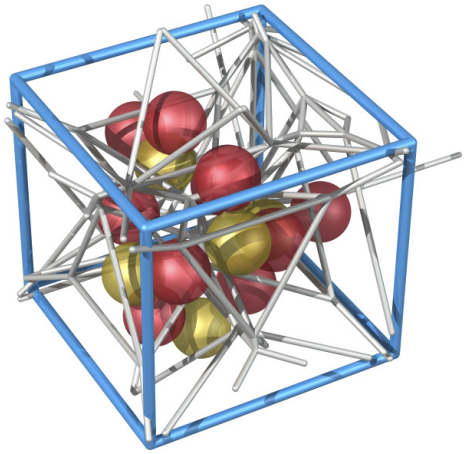
[44] N.E.R. Zimmermann, M. Haranczyk, M. Sharma, B. Liu, B. Smit, F.J. Keil, *Molecular Simulations*, in press.

[45] M. Haranczyk, J.A. Sethian, Proc. Natl. Acad. Sci. USA 106 (2009) 21472-21477.

Figures and Figure captions

Figure 1: Example calculation of the Voronoi networks for the (a) EDI zeolite in a rectangular unit cell, and (b) NPO zeolite in a non-rectangular unit cell. Oxygen and silicon atoms are shown in red and yellow respectively. The unit cells are shown in blue, and the computed Voronoi networks are shown in white. The edges of the Voronoi networks that protrude from the unit cell connect to neighboring periodic unit cells (not shown for clarity of the illustration).

(a)



(b)

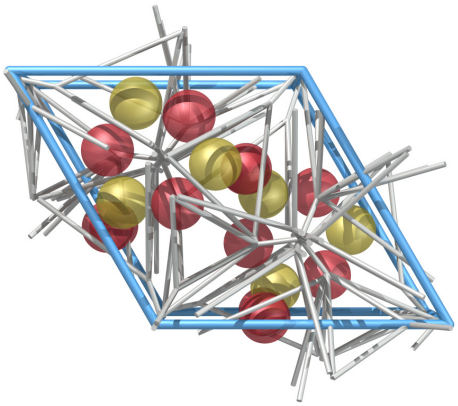


Figure 2. Sampled points on the surface of DDR. Green and red points are respectively, accessible and inaccessible to a spherical probe of radius 3.2Å.

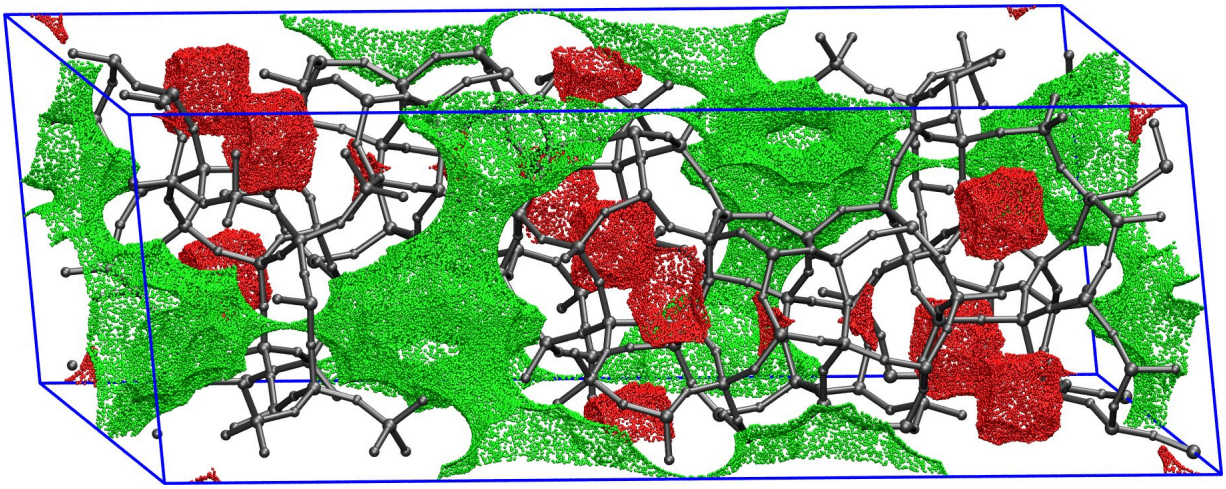
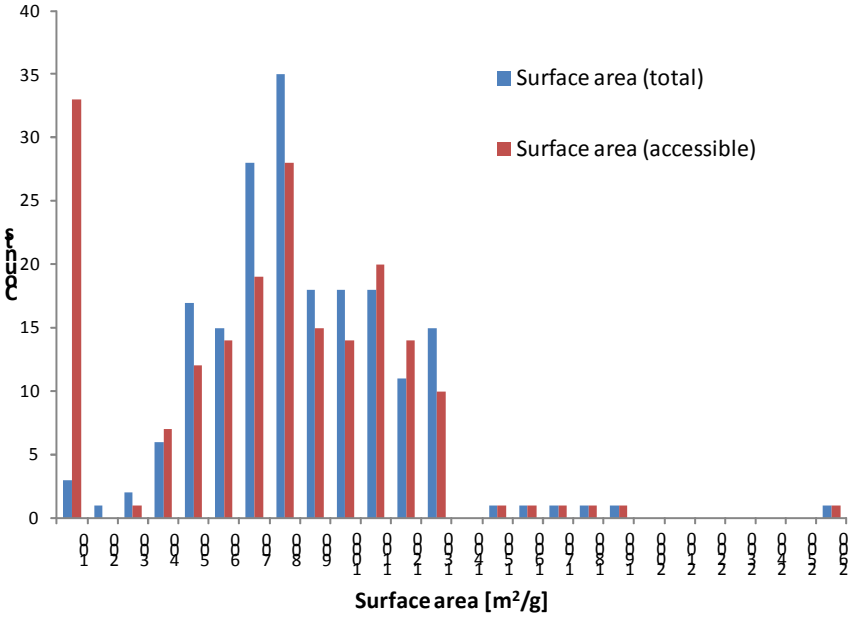


Figure 3. Histograms of surface area and void volume for a set of 193 IZA zeolites. Surface area and void volume were calculated using spherical probes of radii of 1.6 Å and 1.4Å, respectively. Quantities denoted as accessible exclude the contributions corresponding to probe-inaccessible pockets.



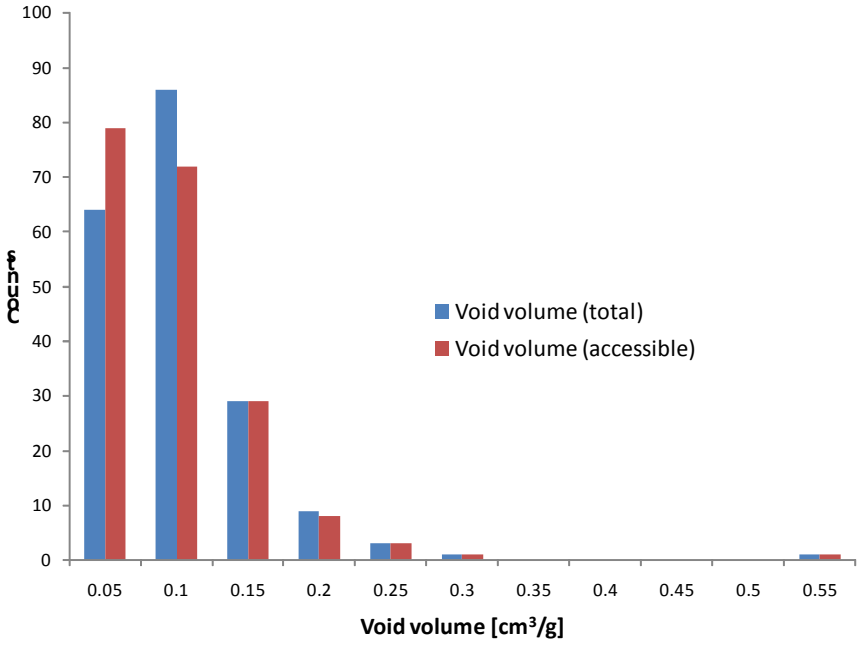


Figure 4. Pore landscapes for four zeolites. Internal surface of pores highlighted in orange.

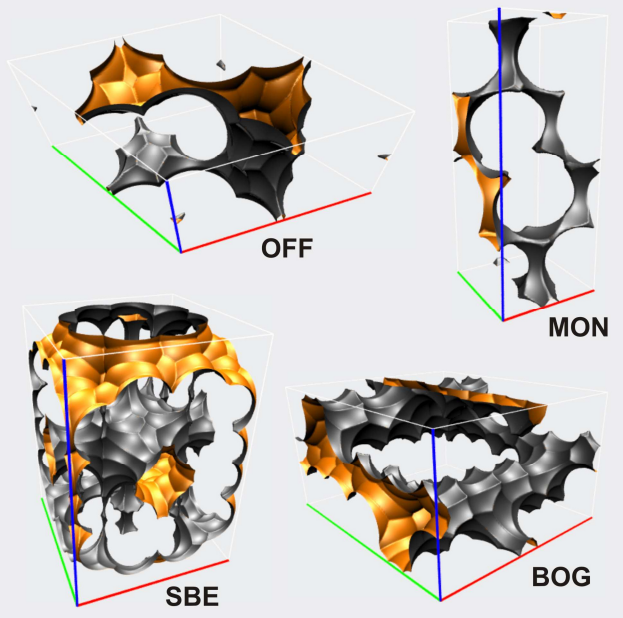


Table 1. Maximum included sphere diameters, D_i, and largest free sphere diameter d_f for the 29 known zeolite frameworks not listed in Ref. 17.

Framework type	D_i	d_f
BOF	5.52	4.61
BSV	5.11	3.78
EZT	6.51	6.07
FAR	6.30	2.41
IHW	6.61	3.61
IMF	7.28	5.38
ITR	6.30	5.06
IWS	8.19	6.60
IWV	8.48	6.97
JRY	4.53	4.34
LTF	8.10	7.44
MOZ	9.97	7.48
MRE	6.30	5.53
MSE	7.03	6.53
MVY	3.70	2.88
PUN	5.45	4.29
SAF	6.60	6.13
SBN	5.00	3.74
SFS	7.46	5.86
SIV	5.32	3.67
SOF	5.08	4.36
SSF	7.60	6.16
STO	6.74	6.03
STW	5.37	4.82
SVR	5.79	5.03
SZR	6.21	4.63
TOL	6.31	2.30
TUN	8.40	5.33
UOS	5.79	4.18